

Naive Bayes

Problem

Implement a spam filter based on the Naive Bayes classifier.

Your classifier should be able to control the number of legit messages labeled as spam. Plot the dependence of classifier accuracy on the number of legit messages labeled as spam.

File Format

Use [this dataset](#) to test your classifier. This dataset already was splits to 10 parts for cross-validation. Each file contains one message and its name contains class as substring 'spmsg' or 'legit'. Each number in file encode one word.

The first line of a file contains a subject of the message.

The second line is empty.

The third line contains content of the message.